Challenge 9 - Adversarial Perturbation Effect

Author : Tejas Krishna Reddy

Date: 9th December 2020

NUID: 001423166

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df = pd.read_csv('android_traffic.csv')
df.head(3)
```

Out[2]:

	tcp_packets	dist_port_tcp	external_ips	vulume_bytes	udp_packets	tcp_urg_packet	sourc
0	36	6	3	3911	0	0	
1	117	0	9	23514	0	0	
2	196	0	6	24151	0	0	

Data Cleaning:

In [3]:
```
## Checking if there are any missing values - There are no missing values
df.isnull().sum()
```

Out[3]:
```
tcp_packets             0
dist_port_tcp           0
external_ips            0
vulume_bytes            0
udp_packets             0
tcp_urg_packet          0
source_app_packets      0
remote_app_packets      0
source_app_bytes        0
remote_app_bytes        0
source_app_packets.1    0
dns_query_times         0
type                    0
dtype: int64
```

In [4]:
```
## Check for the type of classes
df.type.value_counts()
```

Out[4]:
```
benign       4704
malicious    3141
Name: type, dtype: int64
```

In [5]:
```
### Convert the labels to 0's and 1's
df['type'] = df.type.map(dict(benign = 1, malicious = 0))
df.type
```

Out[5]:
```
0       1
1       1
2       1
3       1
4       1
       ..
7840    0
7841    0
7842    0
7843    0
7844    0
Name: type, Length: 7845, dtype: int64
```

Data preprocessing:

- Scaling

In [6]:
```
X = df.drop(['type'],1)
y = df['type']
```

In [7]:
```
from sklearn.preprocessing import StandardScaler
cols = list(X.columns)
scaler = StandardScaler()
X = scaler.fit_transform(X)
X = pd.DataFrame(X)
X.columns = cols
```

In [8]:
```
X.head(2)
```

Out[8]:

	tcp_packets	dist_port_tcp	external_ips	vulume_bytes	udp_packets	tcp_urg_packet	sourc
0	-0.143441	-0.033652	0.086046	-0.153587	-0.040693	-0.015969	
1	-0.039311	-0.149817	2.138859	0.084743	-0.040693	-0.015969	

In [9]:
```
from sklearn import model_selection

X_train, X_test, y_train, y_test = model_selection.train_test_split(X, y, t
print(X_train.shape, y_train.shape)
print(X_test.shape, y_test.shape)
```

```
(6276, 12) (6276,)
(1569, 12) (1569,)
```

Modelling SVM classifier:

In [10]:
```
from sklearn.svm import SVC
modl = SVC(kernel = 'rbf', degree = 3).fit(X_train, y_train)
print("Training Accuracy of the model = ", modl.score(X_train, y_train))
```

```
Training Accuracy of the model =  0.7200446144040791
```

In [11]:
```
print("Testing Accuracy of the model = ", modl.score(X_test, y_test))
```

```
Testing Accuracy of the model =  0.7291268323773104
```

In [ ]:
```

```

Adversarial Robustness Toolbox

In [26]:
```
#!pip install adversarial-robustness-toolbox
```

```
Requirement already satisfied: adversarial-robustness-toolbox in c:\users\t
ejas\appdata\local\continuum\anaconda3\lib\site-packages (1.5.0)
Requirement already satisfied: scipy>=1.4.1 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbo
x) (1.5.4)
Requirement already satisfied: tqdm in c:\users\tejas\appdata\local\continu
um\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (4.43.
0)
Requirement already satisfied: setuptools in c:\users\tejas\appdata\local\c
ontinuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox)
(50.3.2)
Requirement already satisfied: numpy in c:\users\tejas\appdata\local\contin
uum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (1.1
9.4)
Requirement already satisfied: pydub in c:\users\tejas\appdata\local\contin
uum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (0.2
3.1)
Requirement already satisfied: scikit-learn>=0.22.2 in c:\users\tejas\appda
ta\local\continuum\anaconda3\lib\site-packages (from adversarial-robustness
-toolbox) (0.23.2)
Requirement already satisfied: Pillow in c:\users\tejas\appdata\local\conti
nuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (7.
0.0)
Requirement already satisfied: mypy in c:\users\tejas\appdata\local\continu
um\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (0.79
0)
Requirement already satisfied: resampy in c:\users\tejas\appdata\local\cont
inuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (0.
2.2)
Requirement already satisfied: cma in c:\users\tejas\appdata\local\continuu
m\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (3.0.3)
Requirement already satisfied: ffmpeg-python in c:\users\tejas\appdata\loca
l\continuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbo
x) (0.2.0)
Requirement already satisfied: matplotlib in c:\users\tejas\appdata\local\c
ontinuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbox)
(3.2.1)
Requirement already satisfied: six in c:\users\tejas\appdata\local\continuu
m\anaconda3\lib\site-packages (from adversarial-robustness-toolbox) (1.15.
0)
Requirement already satisfied: statsmodels in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from adversarial-robustness-toolbo
x) (0.11.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\tejas\appda
ta\local\continuum\anaconda3\lib\site-packages (from scikit-learn>=0.22.2->
adversarial-robustness-toolbox) (2.1.0)
Requirement already satisfied: joblib>=0.11 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from scikit-learn>=0.22.2->adversar
ial-robustness-toolbox) (0.13.2)
Requirement already satisfied: mypy-extensions<0.5.0,>=0.4.3 in c:\users\te
jas\appdata\local\continuum\anaconda3\lib\site-packages (from mypy->adversa
rial-robustness-toolbox) (0.4.3)
Requirement already satisfied: typed-ast<1.5.0,>=1.4.0 in c:\users\tejas\ap
```

```
pdata\local\continuum\anaconda3\lib\site-packages (from mypy->adversarial-r
obustness-toolbox) (1.4.1)
Requirement already satisfied: typing-extensions>=3.7.4 in c:\users\tejas\a
ppdata\local\continuum\anaconda3\lib\site-packages (from mypy->adversarial-
robustness-toolbox) (3.7.4.2)
Requirement already satisfied: numba>=0.32 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from resampy->adversarial-robustnes
s-toolbox) (0.52.0)
Requirement already satisfied: future in c:\users\tejas\appdata\local\conti
nuum\anaconda3\lib\site-packages (from ffmpeg-python->adversarial-robustnes
s-toolbox) (0.18.2)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in
c:\users\tejas\appdata\local\continuum\anaconda3\lib\site-packages (from ma
tplotlib->adversarial-robustness-toolbox) (2.4.6)
Requirement already satisfied: cycler>=0.10 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from matplotlib->adversarial-robust
ness-toolbox) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\tejas\appdata
\local\continuum\anaconda3\lib\site-packages (from matplotlib->adversarial-
robustness-toolbox) (1.1.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\tejas\appda
ta\local\continuum\anaconda3\lib\site-packages (from matplotlib->adversaria
l-robustness-toolbox) (2.8.1)
Requirement already satisfied: patsy>=0.5 in c:\users\tejas\appdata\local\c
ontinuum\anaconda3\lib\site-packages (from statsmodels->adversarial-robustn
ess-toolbox) (0.5.1)
Requirement already satisfied: pandas>=0.21 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from statsmodels->adversarial-robus
tness-toolbox) (1.0.5)
Requirement already satisfied: llvmlite<0.36,>=0.35.0 in c:\users\tejas\app
data\local\continuum\anaconda3\lib\site-packages (from numba>=0.32->resampy
->adversarial-robustness-toolbox) (0.35.0)
Requirement already satisfied: pytz>=2017.2 in c:\users\tejas\appdata\local
\continuum\anaconda3\lib\site-packages (from pandas>=0.21->statsmodels->adv
ersarial-robustness-toolbox) (2020.4)
```

In [12]:
```
from art.attacks.evasion import UniversalPerturbation
```

In [19]:
```
from art.estimators.classification import SklearnClassifier
```

In [21]:
```
clf = SklearnClassifier(modl)
```

In [ ]:
```
# Generate adversarial test examples
Uperturb = UniversalPerturbation(classifier = clf, attacker = 'deepfool',
                        delta = 0.2, max_iter = 20, eps = 10.0,
                        batch_size = 32,
                        verbose = True)
```

Generate adversarial test samples from original test data

In [23]:
```
x_test_adv = Uperturb.generate(x=X_test)
```

```
| 1/1 [00:00<00:00,  1.81it/s]

DeepFool:   0%|
| 0/1 [00:00<?, ?it/s]
DeepFool: 100%|
| 1/1 [00:00<00:00,  3.40it/s]

DeepFool:   0%|
| 0/1 [00:00<?, ?it/s]
DeepFool: 100%|
| 1/1 [00:02<00:00,  2.15s/it]

DeepFool:   0%|
| 0/1 [00:00<?, ?it/s]
DeepFool: 100%|
| 1/1 [00:03<00:00,  3.59s/it]

DeepFool:   0%|
| 0/1 [00:00<?, ?it/s]
DeepFool: 100%|
```

In [27]:
```
# Step 7: Evaluate the ART classifier on adversarial test examples

predictions = modl.predict(x_test_adv)
print("Accuracy After Perturbation = ", modl.score(x_test_adv, y_test))
```

```
Accuracy After Perturbation =  0.29509241555130655
```

In [25]:
```
x_test_adv
```

Out[25]:
```
array([[-1.01383992, -2.9713307 , -2.94384061, ...,  0.30660187,
        -0.88522743,  2.87558493],
       [-1.22210081, -3.20365953, -3.97024718, ..., -0.0152251 ,
        -1.09447409,  2.82267284],
       [-0.91485172, -3.20365953, -2.25956957, ...,  0.31312071,
        -0.77996224,  3.14014537],
       ...,
       [-1.12054149, -3.20365953, -4.3123827 , ...,  0.01001267,
        -0.99434378,  2.76976076],
       [-1.25938208, -3.20365953, -4.99665374, ..., -0.10313845,
        -1.13426946,  2.71684867],
       [-1.25938208, -3.20365953, -4.99665374, ..., -0.10313845,
        -1.13426946,  2.71684867]])
```

In [ ]:
```

```

In [ ]:

In [ ]: